AD-A049 296 DEFENCE AND CIVIL INST OF ENVIRONMENTAL MEDICINE DOW--ETC F/G 12/1
A TECHNIQUE FOR CALCULATING THE PARAMETERS OF A NORMAL OR LOGNO--ETC(U)
OCT 76 R Y NISHI, P B GIRY
UNCLASSIFIED DCIEM-TR-76-X-71 NL

1 of 1
ADA049 296

END
DATE
FILMED
3 -78
DDC

(3)

(11) OCTOBER 1976

(12) 31p.

(14) DCIEM Technical Report, No. 76-X-71

(9)

TR-76-X-71

(6) A TECHNIQUE FOR CALCULATING THE PARAMETERS OF A NORMAL OR LOGNORMAL CUMULATIVE DISTRIBUTION.

(10) R.Y. Nishi

P.B.L. Giry

Biosciences Division
Defence and Civil Institute of Environmental Medicine
1133 Sheppard Avenue West, P.O. Box 2000
Downsview, Ontario M3M 3B9

DEPARTMENT OF NATIONAL DEFENCE - CANADA

406 986

LB

TABLE OF CONTENTS

ABSTRACT

The statistical distribution of biological phenomena is generally assumed to be normal or Gaussian. In some instances, however, the distribution is lognormal, that is, the logarithm of the variable is normally distributed. This report presents a technique for determining the mean and standard deviation of these distributions from the cumulative distribution function. The actual distribution is compared against a standardized cumulative distribution function of mean, 5, and standard deviation, 1, (probit transformation). The relationship between the random variable and the probit is found by linear regression and the corresponding mean and standard deviation determined. Two programs are presented for solving this problem, one written for a Hewlett-Packard 9820A programmable calculator with plotter, and the other written for a digital computer in Fortran IV. As an example of the use of these programs, the platelet size distribution in a fresh blood sample, obtained from rats subjected to hyperbaric exposures and subsequently decompressed, is solved.

INTRODUCTION

The statistical distribution of biological phenomena is generally assumed to be a normal or Gaussian distribution. In certain instances, however, the distribution can be asymmetric with the logarithm of the variable obeying the normal law of probability. This distribution is commonly described as a lognormal distribution. An example is the platelet volume distribution[6] in a fresh blood sample. Many other examples exist in a variety of fields, most notably in economic data where the lognormal distribution is the rule rather than the exception[3].

The basic problem when one has a normal or lognormal distribution is to obtain the parameters for the distribution, that is, the mean and the standard deviation. If the information is presented as the non-cumulative probability, and a great number of observations are available, it is possible to fit the distribution by standard techniques such as the least squares method for finding the mean and the standard deviation. However, if the information is presented as the cumulative probability, that is, the probability that the observation is less than or equal to some value, the problem becomes more difficult to solve. The cumulative distribution function is a sigmoid curve. If a sufficient number of observations is used to specify this curve, it is possible to extract a noncumulative distribution and use standard techniques. If the number of observations is not large, then one must resort to other methods.

In this paper, a technique for determining the mean and standard deviation from a cumulative distribution function is presented where the actual distribution is compared to the standardized cumulative distrubution function. This technique can be easily programmed on programmable calculators or on a digital computer, and a calculator program and a Fortran IV listing are presented. The same principle has been used previously by Servantie et al[7] but the version presented here is designed for much faster operation with less calculations.

METHOD OF CALCULATION

The general equation for the noncumulative probability density function (p.d.f.) is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad (1)$$

where x is the measured variable

μ is the mean, and

σ is the standard deviation.

The cumulative distribution function (c.d.f.) is given by

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt \qquad (2)$$

where F is the probability that the random variable is less than or equal to x. The standardized p.d.f. is defined for $\mu = 0$ and $\sigma = 1$ and tables giving both the p.d.f. and c.d.f. are readily available.

In the case of the lognormal distribution, the logarithm of the variable x is normally distributed. The p.d.f. is given by

$$f = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2} \quad (x>0 \text{ only}) \qquad (3)$$

and the c.d.f. by

$$F = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\ln x} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt \qquad (4)$$

The simplest method to determine the mean and the standard deviation from the c.d.f. is to use normal or lognormal probability paper, depending on the distribution. The c.d.f. will plot as a straight line and it is a simple matter to determine the mean at the 50% probability point and the standard deviation from the 84.1% probability point. If better accuracy is required or if the observations have sufficient scatter that a linear regression must be carried out on the data, the graphical analysis is inadequate.

The method presented here compares the observed c.d.f. against the standardized c.d.f. The probit transformation

$$P = x + 5 \qquad (5)$$

is used to avoid negative values so that the standardized c.d.f. becomes

$$F = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{P} e^{-\frac{1}{2}(x - 5)^2} dx \qquad (6)$$

with the mean at 5. For every x at which a probability F is observed, there is a corresponding value, P, from the standardized c.d.f. (equation 6) for the same value of F. If the observed distribution is normal, the relationship between x and P will be linear.

$$P = ax + b \qquad (7)$$

and if the distribution is lognormal

$$P = a \ln x + b. \qquad (8)$$

The slope, a, and the intercept, b, can be obtained by linear regression from

$$a = \frac{n\Sigma PX - \Sigma P\Sigma X}{n X^2 - (\Sigma X)^2} \qquad (9)$$

$$b = \frac{\Sigma P - a\Sigma X}{n} \qquad (10)$$

where n is the total number of observations and X is either x or lnx depending on the distribution.

The mean, m, can be found from P = 5, hence

$$m = (5 - b)/a \qquad (11)$$

and the standard deviation, s, from

$$s = 1/a. \qquad (12)$$

For the lognormal distribution[5], it is useful to have also the mode

$$\text{mode} = e^{(m-s^2)} \tag{13}$$

the median

$$\text{median} = e^{m} \tag{14}$$

and the expectation value

$$m_d = e^{(m + 0.5s^2)} \tag{15}$$

Although the method is straightforward, the problem lies in determining P for a given value of F (from equation 6). Since the probability integral can only be solved numerically, i.e. F as a function of P, the inverse problem, P as a function of F, is difficult to solve. Servantie et al[7] use an iterative process with

$$\frac{1}{\sqrt{2\pi}} \Sigma e^{-x^2/2} \Delta x$$

by subdividing Δx until the calculated F is sufficiently close to the observed F and then determining the value of x. The process is, however, extremely slow for use on programmable calculators.

The technique shown here is based on a rational approximation to the inverse probability integral. If the probability integral is written as

$$Y = \frac{1}{\sqrt{2\pi}} \int_{x_p}^{\infty} e^{-x^2/2} dx \tag{16}$$

with $0<Y\leq 0.5$, then

$$x_p = t - \frac{C_0 + C_1 t + C_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} + \varepsilon(Y) \tag{17}$$

where $|\varepsilon(Y)| < 4.5 \times 10^{-4}$

and $t = \sqrt{\ln(1/Y^2)}$ (18)

The set of constants, C and d, are given by

$C_0 = 2.515517$ $d_1 = 1.432788$

$C_1 = 0.802853$ $d_2 = 0.189269$

$C_2 = 0.010328$ $d_3 = 0.001308$

The probit, P, can then be determined from equations 16 and 17 for the following ranges of the observed values, F,

$0<F\leq 0.5,$ $Y = F,$ $P = 5 - x_p$

$0.5<F<1.0,$ $Y = 1 - F,$ $P = 5 + x_p$

EXAMPLE

The technique presented here has been used to find the mean and standard deviation of platelet sizes in blood samples obtained from rats which were subjected to hyperbaric exposures and subsequently decompressed[4]. The data have been obtained on a Coulter Counter and Channelizer. Table I shows the data obtained from six samples of blood.

Programmable calculator solution - Appendix I contains a listing of a program written for a Hewlett-Packard 9820A Calculator with plotting facilities. Appendix II gives instructions on its use, and a flow chart is shown in Figure 1. Data points (up to 30 observations) are input from the keyboard, the probabilities and probits calculated, and stored in memory. Provisions are made for correcting errors after all the data have been entered. Either a normal analysis or lognormal analysis can be selected from the keyboard. A plot of the data and calculated fit to the data is optional.

Appendix III shows a listing of the input data and the calculated results for a lognormal analysis. The data are plotted in Figure 2 and it can be observed that the probit values (indicated by "+") when plotted against ln x are linear except for the last four observations. This departure from linearity results since the Coulter can not differentiate between large platelets and small red blood cells which are present in this range. The program has provisions for omitting these observations from the analysis. The results in Appendix III are based on the first 15 observations only. The solid lines in Figure 2 show

the values calculated for the probit and the c.d.f. using the computed parameters a, b, m and s. The c.d.f. is calculated by an iterative method. The results indicate that the distribution of platelet sizes is indeed lognormal.

The program has provisions for repeating the analysis with a different number of omitted observations, if for example, a better fit is required for the experimental data. If a repeat is not required, then the program branches back to the location for selecting either the normal or lognormal analysis again, or for terminating the program. Appendix III also shows results for a normal distribution analysis. Figure 3 shows that the platelet sizes are not normally distributed.

Digital computer solution (FORTRAN IV) - Appendix IV shows a Fortran IV listing for use on a digital computer. The data is prepared as an input file on a peripheral device in the format shown in Appendix V. Up to 100 observations can be used, although this can be easily changed by modifying the program. The program is an interactive one where the operator can choose a particular input data file and normal or lognormal analysis from the teletype. If a line printer is used for output, then this must be signified.

Appendix VI shows the results (teletype output) for the lognormal analysis of the platelet data. Initially, the parameters are calculated using all the observations and printed out on the output device. The probit, calculated from the computed values of the slope, a, and intercept, b, is printed out as the difference between the calculated and observed values, since a plotter is not used. These differences and the correlation coefficient can be used to tell whether the analysis selected is appropriate (i.e., normal or lognormal). If there is some doubt, then the analysis can be repeated with the other type of distribution and compared. The correct analysis will give a larger correlation coefficient.

The differences between the calculated and the observed values at the tails of the distribution can be examined to determine whether any observation should be discarded. The program asks for the number of observations to be omitted at the beginning and end of the data file. The program then calculates new parameters and the probit differences. The analysis can be repeated with different numbers of observations to be discarded, if desired.

When the analysis is to be terminated, the observed data, the c.d.f., the probit values, and the differences between the observed c.d.f. and those calculated from the mean and standard deviation are printed out. The c.d.f. calculated from the mean and standard deviation uses the rational approximation[2]

$$Y(x) = 1 - Z(x)(a_1 t + a_2 t^2 + a_3 t^3) + \varepsilon(x)$$

where $0 \leq x < \infty$

$$t = 1/(1+px)$$

$$|\varepsilon(x)| < 1 \times 10^{-5}$$

$$Z(x) = \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}}$$

$$Y(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}w^2} dw$$

and $p = 0.33267$

$a_1 = 0.4361836$

$a_2 = -0.1201676$

$a_3 = 0.937298$

The agreement between the calculated c.d.f. and the observed c.d.f. is excellent. After the printout is completed, the program will request a new input file.

Appendix VII shows the same data treated as a normal distribution (line printer output). With the line printer output, the correlation coefficient and probit differences are also printed on the teletype. The correlation coefficient is less than that for the lognormal distribution indicating that the normal distribution is a poor fit to the observed distribution. This is borne out by the large errors between the calculated and observed values of the probit and the poor agreement between the observed and calculated probabilities.

DISCUSSION AND SUMMARY

Of the two methods presented here, the programmable calculator method is probably the most useful for most applications since a graphical plot is available and it is easy to see which data points are spurious and how well the calculated values agree with the experimental observations. It is limited in the amount of data that can be input because of the size of the memory available in the calculator

for data storage. The program can be easily rewritten to calculate the various sums continuously so that the input data need not be stored. Hence an unlimited number of observations can be used. However, it would not be convenient to omit data points which may be spurious as in the example where the red blood cells inflate the observed platelet count. With the present program, also, the data can be saved on a magnetic card for future use if necessary.

The digital computer program is better suited for analyzing samples where a large number of observations are available since the calculations can be done much faster. If a large number of samples are to be analyzed, all the samples can be prepared beforehand as separate input files and edited for mistakes before running them through the program. The data will also be available for reanalysis at a later time if necessary without having to key in all the data again. A disadvantage of the program is the lack of a graphical indication of the observed values and the calculated values. The program can be easily modified if a digital XY plotter is available.

Both programs described here provide a much more convenient method for finding the mean and standard deviation from a cumulative probability distribution than other standard techniques such as graphical methods. The use of rational approximations for calculating the probability integral and its inverse provide sufficient accuracy for the type of data generally available, and also allows the calculations to proceed much faster than with iterative numerical methods. Although the example shown here is for the analysis of platelet sizes in a fresh blood sample, the technique can be applied to any sort of data which may be normally or lognormally distributed.

REFERENCES

1 ABRAMOWITZ, M. & I.A. STEGUN, Ed. Handbook of mathematical functions, National Bureau of Standards, Applied Mathematics Series 55, 1966, p. 933, 26.2.23.

2 Ibid., p. 932, 26.2.16.

3 AITCHISON, J. & J.A.C. BROWN. The lognormal distribution, Cambridge University Press, London, 1957.

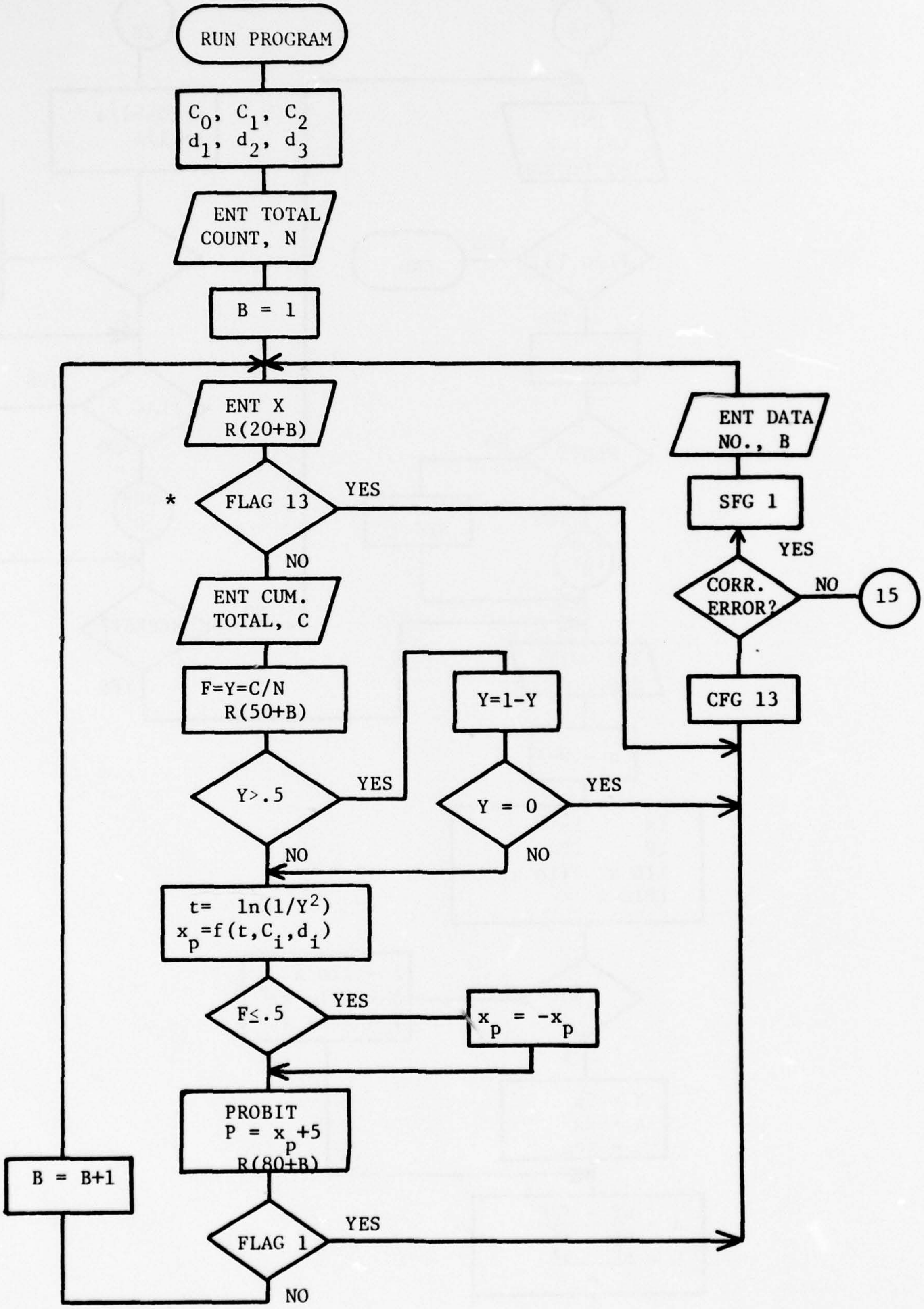
4 GIRY P., G. PORLIER, D. EASTMAN, & M.W. RADOMSKI. Dive-induced modifications in platelet kinetics in rats. DCIEM Publication No. 76-X-62, 1976.

5 MEYER S.L. Data analysis for scientists and engineers, John Wiley and Sons, New York, 1975, p. 285.

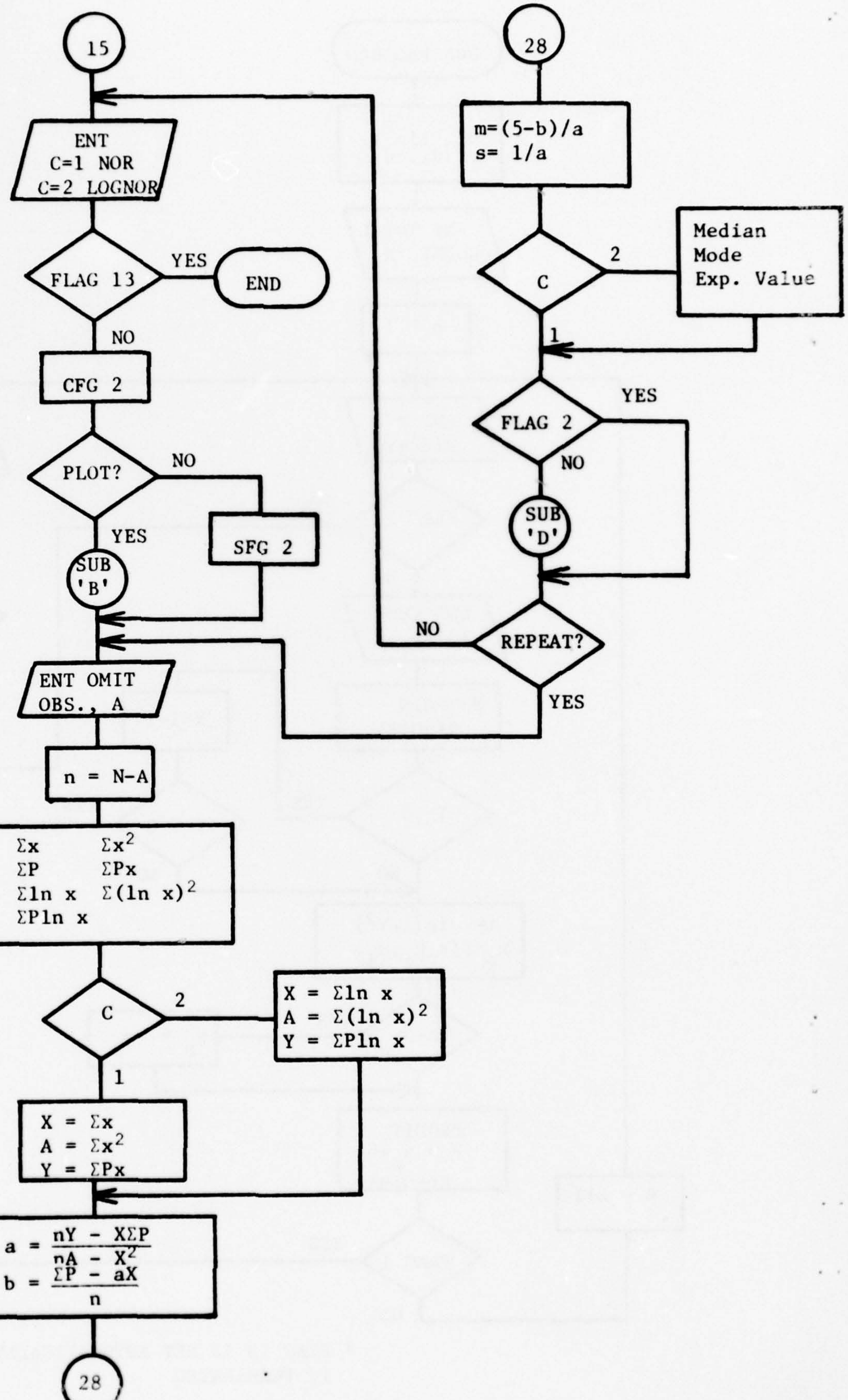
6 PAULUS J.M. Production et destruction des plaquettes sanguines. These Pour L'obtention du Titre D'agrégé de L'enseignement Superieur, Liege (Belgium), 1972 Masson, Paris.

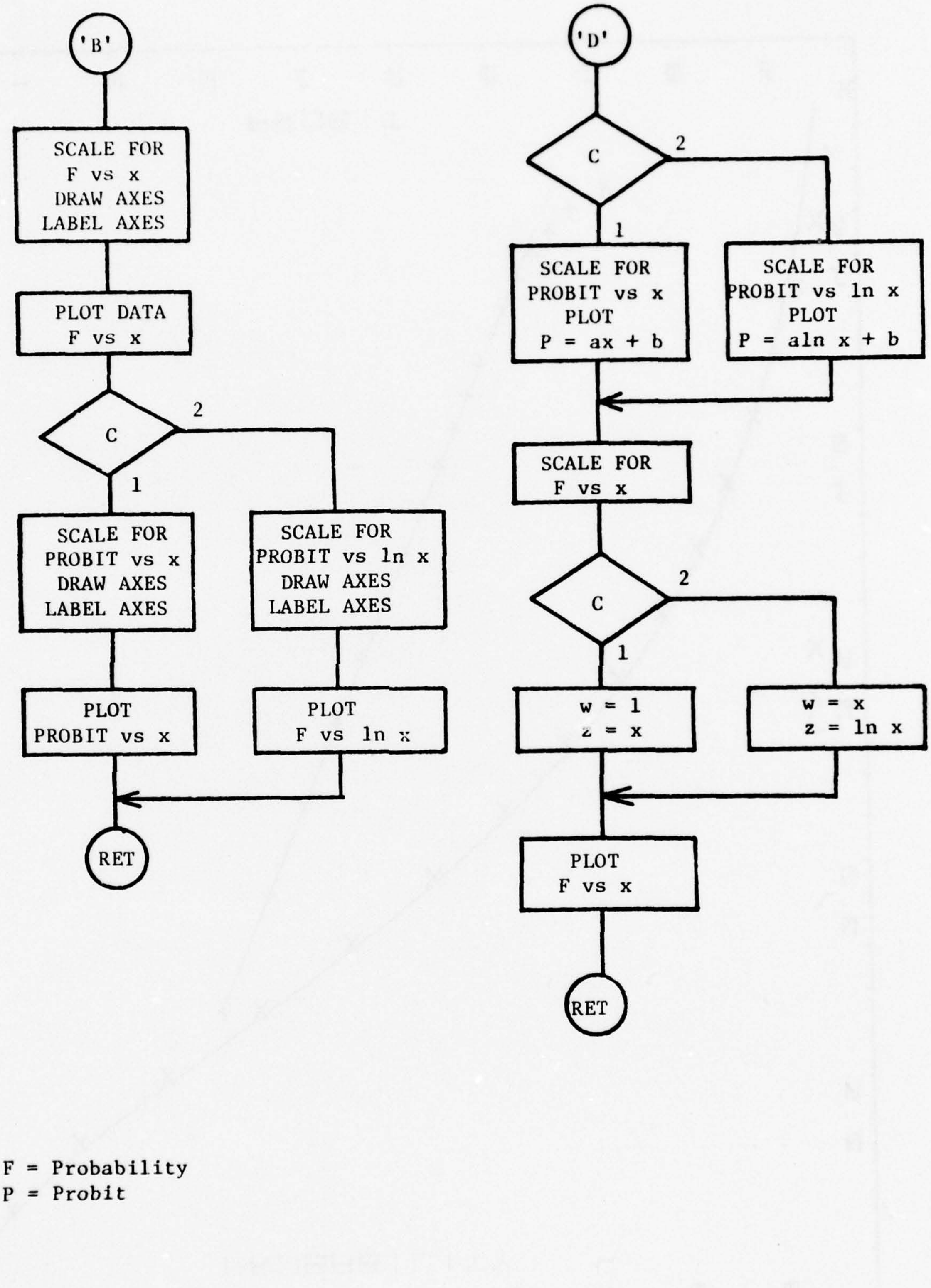
7 SERVANTIE B., P.B.L. GIRY, T. OBRENOVITCH, & F. BRUE. Exploitation statistique des distributions sigmoidales par calculateur programmable, *Centre D'études et de Recherches Bio-*Physiologique Appliquées a la Marine, Toulon, France, Rapport No. 75-20, 1975.

Figure 1: Flow chart for calculator program



* FLAG 13 IS SET AUTOMATICALLY WHEN INPUT IS TERMINATED


15
ENT
C=1 NOR
C=2 LOGNOR
FLAG 13
YES
END
NO
CFG 2
PLOT?
NO
YES
SFG 2
SUB 'B'
ENT OMIT OBS., A
n = N-A
Σx Σx²
ΣP ΣPx
Σln x Σ(ln x)²
ΣPln x
C
2
X = Σln x
A = Σ(ln x)²
Y = ΣPln x
1
X = Σx
A = Σx²
Y = ΣPx
a = (nY - XΣP)/(nA - X²)
b = (ΣP - aX)/n
28
28
m=(5-b)/a
s= 1/a
C
2
Median
Mode
Exp. Value
1
FLAG 2
YES
NO
SUB 'D'
NO
REPEAT?
YES



F = Probability
P = Probit

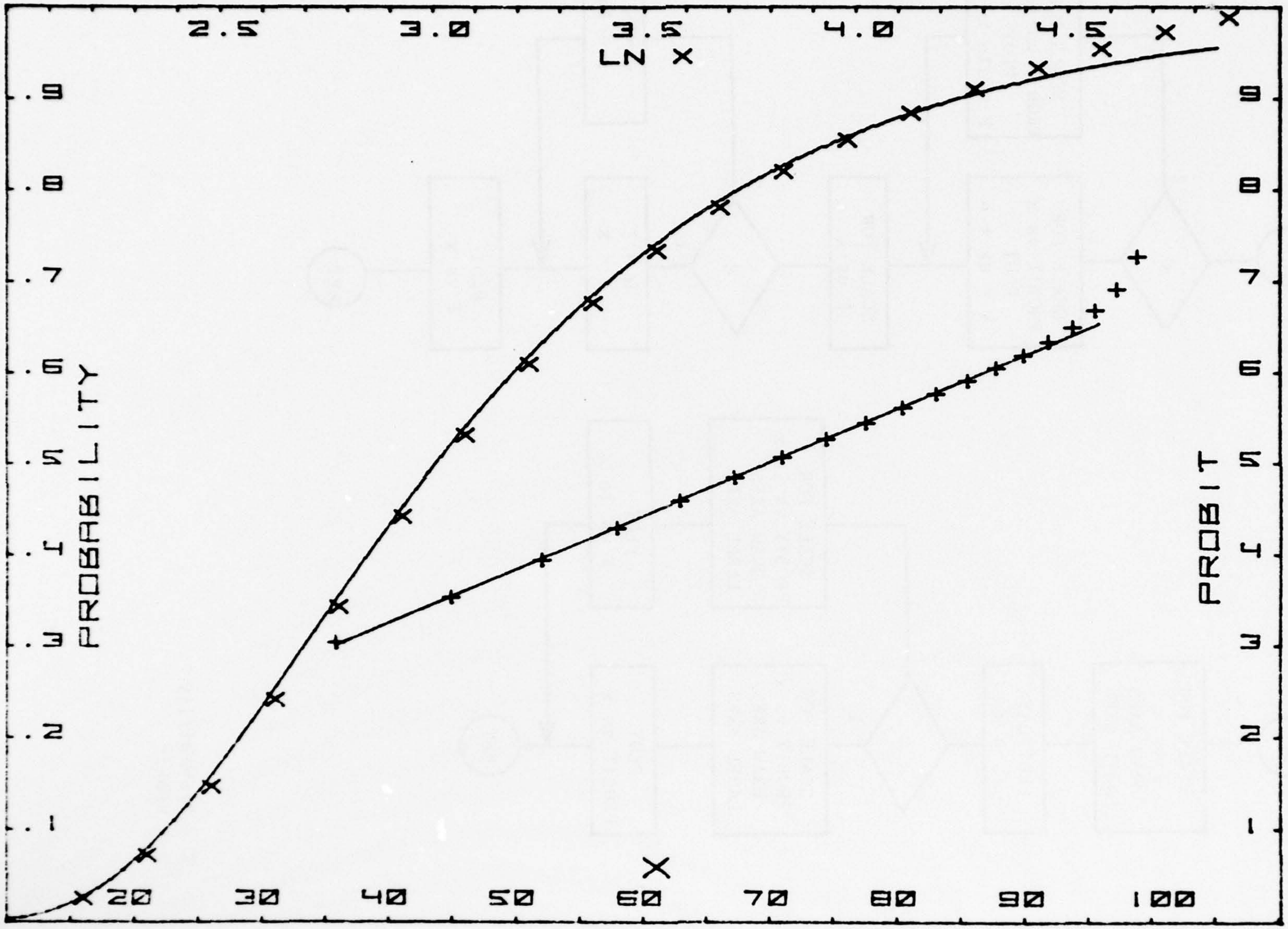


Figure 2: Lognormal analysis of platelet size distribution

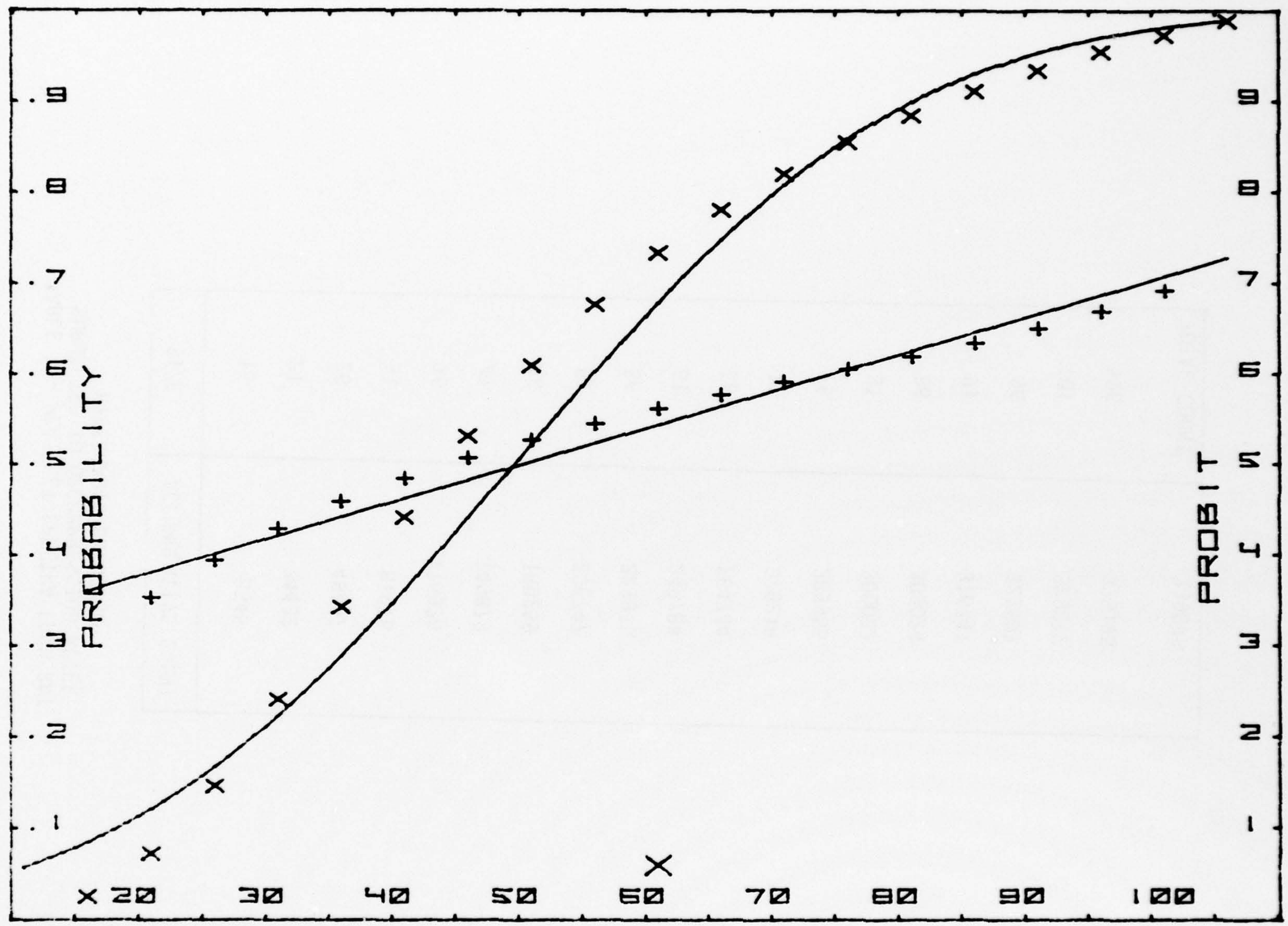


Figure 3: Normal analysis of platelet size distribution

TABLE I - PLATELET SPECTRA FROM RAT SUBJECTED TO DECOMPRESSION AFTER HYPERBARIC EXPOSURE

SIZE	ACCUMULATIVE COUNT
16	8545
21	24432
26	49667
31	81915
36	116328
41	149879
46	180259
51	206747
56	229375
61	248784
66	264744
71	278218
76	289956
81	300032
86	309024
91	316741
96	323697
101	329812
106	335196
TOTAL COUNT	339075

APPENDIX I Program listing for HP 9820A calculator


```
0:
TBL 4;2.515517→R
0;.802853→R1;.01
0328→R2⊢
1:
1.432788→R11;.18
9269→R12;.001308
→R13;CFG 13;FXD
0;CFG 1⊢
2:
ENT "TOTAL COUNT
",R20;PRT "TOTAL
 COUNT",R20;1→B;
SPC 2⊢
3:
PRT "DATA NO.","
X-VALUE","CUM. T
OTAL","C.D.F.","
PROBIT"⊢
4:
FXD 0;PRT B;ENT
X;IF FLG 13;GTO
13⊢
5:
X→R(20+B);ENT "C
UM. TOTAL",C;
FXD 1⊢
6:
PRT X,C;FXD 4;
PRT C/R20→R(50+B
)→C⊢
7:
C→X;IF C>.5;1-C→
X;IF X=0;GTO 13⊢
8:
√LN (1/XX)→A⊢
9:
A-(R0+A(R1+AR2))
/(1+A(R11+A(R12+
AR13)))→A;IF C≤.
5;-A→A⊢
10:
A+5→R(80+B)→A;
PRT A;SPC ⊢
11:
IF FLG 1;GTO +2⊢
12:
B→R19;B+1→B;GTO
4⊢
13:
CFG 13;ENT "CORR
ECT ERROR?",A;
IF FLG 13;GTO +2
⊢
14:
SFG 1;ENT "DATA
NO.",B;PRT "****
************";
GTO 4⊢
15:
CFG 13;ENT "NOR=
1,LOG=2",C;IF
FLG 13;GTO 68⊢
16:
CFG 1;CFG 2;ENT
"PLOT ?",A;IF
FLG 13;SFG 2;
GTO 18⊢
17:
GSB "B"⊢
18:
CFG 13;0→A;ENT "
OMIT ?OBS.AT END
",A;R19-A→R3⊢
19:
CFG 13;1→B;SPC 3
;0→R4→R5→R6→R7→R
8→R9→R10⊢
20:
R(20+B)→X;LN X→Z
;R(80+B)→Y⊢
21:
R4+X→R4;R5+XX→R5
;R8+Z→R8;R9+ZZ→R
9⊢
22:
R6+Y→R6;R7+XY→R7
;R10+ZY→R10⊢
23:
B+1→B;IF B≤R3;
GTO -3⊢
24:
IF C=1;R4→X;R5→A
;R7→Y;PRT "NORMA
L";SPC ⊢
25:
IF C=2;R8→X;R9→A
;R10→Y;PRT "LOG
NORMAL";SPC ⊢
26:
(R3Y-XR6)/(R3A-X
X)→A→R14⊢
27:
(R6-AX)/R3→B→R15
⊢
28:
FXD 4;PRT "P=AX+
B","A",A,"B",B⊢
29:
(5-B)/A→X→R16;1/
A→Y→R17⊢
30:
PRT "MEAN--PROBI
T=5",X;IF C=2;
PRT "MEDIAN",
EXP X⊢
31:
PRT "ST. DEV.",Y
;IF C=2;PRT "MOD
E",EXP (X-YY),"E
XP. VALUE",EXP (
X+.5YY)⊢
32:
SPC ;FXD 0;PRT "
NO. OBS. USED",R
3;SPC 8;IF FLG 2
;GTO +3⊢
33:
GSB "SCL 2"⊢
34:
GSB "D"⊢
35:
ENT "REPEAT ?",A
;IF FLG 13;GTO 1
5⊢
36:
GTO 18⊢
37:
"B";CFG 13;GSB "
SCL 1"⊢
38:
AXE 10,0,5,.1;.1
→B;FXD 1⊢
39:
LTR 11,B,221;
PLT 8;.1+B→B;IF
B≤.9;GTO +0⊢
```

```
40:
LTR 17,.3,232;
PLT "PROBABILITY
";20→B;FXD 0⊢
41:
LTR B-2,.02,221;
PLT B;10+B→B;IF
B≤100;GTO +0⊢
42:
LTR 60,.05,331;
PLT "X";1→B⊢
43:
LTR R(20+B)-.64,
R(50+B)-.0064,22
1;PLT "X";B+1→B;
IF B≤R19;GTO +0⊢
44:
GSB "SCL 2"⊢
45:
IF C=2;AXE 5,10,
.1,1;.0192→A;1→B
;GTO +2⊢
46:
AXE 110,10,5,1;.
64→A;1→B⊢
47:
LTR Z-.03(Z-X),B
,221;PLT B;1+B→B
;IF B≤9;GTO +0⊢
48:
LTR Z-.05(Z-X),3
.5,232;PLT "PROB
IT";2.5→B⊢
49:
IF C=2;FXD 1;
LTR B-.07,9.7,22
1;PLT B;.5+B→B;
IF B≤4.5;GTO +0⊢
50:
IF C=2;LTR 3.4,9
.4;PLT "LN X";1→
B⊢
51:
R(20+B)→X;IF C=2
;LN R(20+B)→X⊢
52:
LTR X-A,R(80+B)-
.064,221;PLT "+"
;B+1→B;IF B≤R19;
GTO -1⊢
53:
RET ⊢
54:
"D";R21→X;5→A;R(
20+R19)→Y;IF C=2
;LN R21→X;.2→A;
LN R(20+R19)→Y⊢
55:
PLT X,R14X+R15;X
+A→X;IF X≤Y;GTO
+0⊢
56:
PEN ;GSB "SCL 1"
⊢
57:
-50→X;0→R18;IF C
=2;.1→X⊢
58:
1→B;X→A;IF C=2;
LN X→A;X→B⊢
59:
R18+EXP (-.5((A-
R16)/R17)↑2)/BR1
7√(2π)→R18;IF X≤
10;X+1→X;GTO -1⊢
60:
PLT X,R18;X+1→X;
IF X≤R(20+R19);
GTO -2⊢
61:
PEN ⊢
62:
RET ⊢
63:
"SCL 1";SCL 10,1
10,0,1⊢
64:
RET ⊢
65:
"SCL 2";IF C=2;
SCL 2→X,5→Z,0,10
;GTO +2⊢
66:
SCL 10→X,110→Z,0
,10⊢
67:
RET ⊢
68:
END ⊢
R148
```

APPENDIX II Instructions for use of calculator program

STEP	DISPLAY	INSTRUCTIONS
1		ERASE, LOAD, EXECUTE, Load Side 1 EXECUTE, Side 2, EXECUTE, Side 3 EXECUTE, Side 4
2		END, RUN PROGRAM
3	TOTAL COUNT	Enter Total Count, RUN PROGRAM
4	X	Enter x - value, RUN PROGRAM
5	CUM. TOTAL	Enter y-value, RUN PROGRAM
6		Repeat steps 4 and 5 for all data points
7	X	To terminate input, RUN PROGRAM
8	CORRECT ERROR?	To correct error, 1, RUN PROGRAM If no errors, RUN PROGRAM (Go to step 13)
9	DATA No.	Enter data number of incorrect value, RUN PROGRAM
10	X	Correct value of X, RUN PROGRAM
11	CUM. TOTAL	Correct value of y, RUN PROGRAM
12	CORRECT ERROR?	If further errors, go to step 8 If no more errors, RUN PROGRAM
13	NOR = 1, LOG = 2	For normal analysis, 1, RUN PROGRAM For lognormal analysis, 2, RUN PROGRAM
14	PLOT?	If plotter used, 1, RUN PROGRAM If plotter not used, RUN PROGRAM
15	OMIT? OBS. AT END	Enter number of observations to be omitted, RUN PROGRAM If no omission, RUN PROGRAM
16	REPEAT?	To repeat analysis with different number of omitted observations, 1, RUN PROGRAM and go to step 15 To terminate analysis or to try other distribution, RUN PROGRAM
17	NOR = 1, LOG = 2	To repeat on same data, go back to step 13 To terminate program, RUN PROGRAM

APPENDIX III Input data and results for calculator program

BEST AVAILABLE COPY

TO [illegible]
339075

DATA NO.	X-VALUE	CUM. TOTAL	C.D.F.	PROBIT
1	16.0	8545.0	.0252	3.0430
2	21.0	24432.0	.0721	3.5391
3	26.0	49667.0	.1465	3.9483
4	31.0	81915.0	.2416	4.2991
5	36.0	116328.0	.3431	4.5964
6	41.0	149879.0	.4420	4.8545
7	46.0	180259.0	.5316	5.0791
8	51.0	206747.0	.6097	5.2782
9	56.0	229375.0	.6765	5.4574
10	61.0	248784.0	.7337	5.6237
11	66.0	264744.0	.7808	5.7746
12	71.0	278218.0	.8205	5.9173
13	76.0	289956.0	.8551	6.0587
14	81.0	300032.0	.8849	6.1997
15	86.0	309024.0	.9114	6.3495
16	91.0	316741.0	.9341	6.5076
17	96.0	323697.0	.9546	6.6921
18	101.0	329812.0	.9727	6.9222
19	106.0	335196.0	.9886	7.2759
20				

LOG NORMAL

P=AX+B

A
1.9586

B
-2.4187

MEAN--PROBIT=5
3.7877

MEDIAN
44.1525

ST. DEV.
.5106

MODE
34.0211

EXP. VALUE
50.2990

NO. OBS. USED
15

NORMAL

P=AX+B

A
.0408

B
2.9538

MEAN--PROBIT=5
50.1447

ST. DEV.
24.5059

NO. OBS. USED
19

APPENDIX IV Fortran IV listing for PDP-9T computer

```
C     FILE NORMD FOR CALCULATING THE PARAMETERS OF A NORMAL
C     OR LOGNORMAL DISTRIBUTION
C
            LOGICAL FOUND
            DIMENSION FILE(2),TITL(14),X(100),Y(100),XZ(100),PR(100)
            DIMENSION PT(100),DPR(100),DPT(100)
            DATA FILE(2)/4H SRC/
            DATA C0,C1,C2/2.515517,0.802853,0.010328/
            DATA D1,D2,D3/1.432788,0.189269,0.001308/
            DATA A1,A2,A3,AP/.4361836,-.1201676,.937298,.33267/
C
            WRITE (6,300)
300         FORMAT(50H *IF USING LINE PRINTER - TYPE 1 FOR YES, 0 FOR NO)
            READ(4,305) LP
C
1           WRITE(6,301)
301         FORMAT(///21H *TYPE DATA FILE NAME/)
            READ(4,302) FILE(1)
302         FORMAT(A5)
            CALL FSTAT(5,FILE,FOUND)
            IF (FOUND) GO TO 2
            WRITE(6,303)
303         FORMAT(15H FILE NOT FOUND)
            GO TO 1
C
2           WRITE(6,304)
304         FORMAT(/36H *TYPE 1 FOR NORMAL, 0 FOR LOGNORMAL)
            READ(4,305)K
305         FORMAT(I1)
C
            CALL SEEK(5,FILE)
            READ(5,101) TITL
101         FORMAT(14A5)
            WRITE(7,201)TITL
201         FORMAT(1H1,14A5)
            IF (K.EQ.0) GO TO 3
            WRITE(7,202)
202         FORMAT(/29H NORMAL DISTRIBUTION ANALYSIS)
            GO TO 4
3           WRITE(7,203)
203         FORMAT(/32H LOGNORMAL DISTRIBUTION ANALYSIS)
4           READ(5,102) XN
102         FORMAT(F10.2)
            READ(5,102)TOT
            NN=NF=N=IFIX(XN)
            NS=1
            WRITE(7,204) N,TOT
204         FORMAT(/20H NO. OF OBSERVATIONS,5X,I3//
            1 12H TOTAL COUNT,5X,F10.2/)
            IF (K.EQ.0) GO TO 11
            WRITE(7,205)
205         FORMAT(11H P = AX + B)
            GO TO 12
11          WRITE(7,206)
206         FORMAT(15H P = ALN(X) + B)
```

```
C
C    READ X AND Y  AND CALCULATE PROBITS
C
12          DO 41 I=1,N
            READ(5,102) X(I)
            READ(5,102) Y(I)
            IF((TOT-Y(I)).LT..001) GO TO 38
            PR(I)=Y(I)/TOT
            IF (PR(I).LE..5) Q=PR(I)
            IF (PR(I).GT..5) Q=1.-PR(I)
            T=SQRT(ALOG(1./(Q*Q)))
            PTEMP=T-(C0+T*(C1+T*C2))/(1.+T*(D1+T*(D2+D3*T)))
            IF (PR(I).LE..5) PT(I)=5.-PTEMP
            IF (PR(I).GT..5) PT(I)=5.+PTEMP
            GO TO 39
38          NN=NF=N-1
39          IF (K.EQ.0) GO TO 40
            XZ(I)=X(I)
            GO TO 41
40          XZ(I)=ALOG(X(I))
41          CONTINUE
C
C    CALCULATE SUMS
C
43          SMX=SMX2=SMP=SMP2=SMPX=0.
            NUSE=0
            DO 45 I=NS,NF
            NUSE=NUSE+1
            SMX=SMX+XZ(I)
            SMX2=SMX2+XZ(I)*XZ(I)
            SMP=SMP+PT(I)
            SMP2=SMP2+PT(I)*PT(I)
45          SMPX=SMPX+PT(I)*XZ(I)
C
C    CALCULATE PARAMETERS OF DISTRIBUTION
C
            TN=FLOAT(NUSE)
            WRITE(7,207)NUSE
207         FORMAT(////38H NO. OF OBSERVATIONS USED IN ANALYSIS ,I3//)
            A=(TN*SMPX-SMP*SMX)/(TN*SMX2-SMX*SMX)
            B=(SMP-A*SMX)/TN
            W=(TN*SMX2-SMX*SMX)*(TN*SMP2-SMP*SMP)
            R=(TN*SMPX-SMX*SMP)/SQRT(W)
            WRITE(7,208) A,B
208         FORMAT(5H A = ,F10.4,5X,5H B = ,F10.4/)
            XM=(5.-B)/A
            S=1./A
            WRITE(7,209)XM,S
209         FORMAT(7H MEAN =,F10.4/
            1 21H STANDARD DEVIATION =,F10.4/)
            IF (K.EQ.1) GO TO 59
            XMODE=EXP(XM-S*S)
            XMED=EXP(XM)
            XMD=EXP(XM+.5*S*S)
            WRITE(7,210)XMODE,XMED,XMD
210         FORMAT(20H MODE = EXP(M-S*S) = ,F10.4/
            1 18H MEDIAN = EXP(M) = ,F10.4/
```

```
     2 37H EXPECTATION VALUE = EXP(M+0.5S*S) = ,F10.4)
59        DO 61 I=NS,NF
61        DPT(I)=(A*XZ(I)+B)-PT(I)
          WRITE(6,311)
311       FORMAT(/30H DIFFERENCE  (AZ + B) - PROBIT)
          WRITE(6,312) (X(I),DPT(I),I=NS,NF)
312       FORMAT(5(F7.1,F7.3))
          WRITE(6,318) R
318       FORMAT(27H CORRELATION COEFFICIENT = ,F8.4/)
          IF(LP.EQ.0) GO TO 62
          WRITE(7,311)
          WRITE(7,312) (X(I),DPT(I),I=NS,NF)
          WRITE(7,318) R
C
C    OPTION FOR OMITTING OBSERVATIONS
C
62        WRITE(6,313)
313       FORMAT(//46H *OMIT OBSERVATIONS - TYPE 1 FOR YES, 0 FOR NO)
          READ(4,305) M
          IF(M.EQ.0) GO TO 63
          WRITE(6,315)
315       FORMAT(19H *HOW MANY AT START)
          READ(4,316) NS
          NS=NS+1
316       FORMAT(I2)
          WRITE(6,317)
317       FORMAT(17H *HOW MANY AT END)
          READ(4,316) NE
          NF=NN-NE
          GO TO 43
C
C    CALCULATE PROBABILITIES FROM MEAN AND STANDARD DEVIATION
C
63        DO 64 I=1,N
          Z=(XZ(I)-XM)/S
          T=1./(1.+AP*Z)
          IF(Z.LT.0.) T=1./(1.-AP*Z)
          F=EXP(-Z*Z/2.)/SQRT(2.*3.14159)
          P=1.-F*T*(A1+T*(A2+A3*T))
          IF(Z.LT.0.) P=1.-P
64        DPR(I)=P-PR(I)
C
C    PRINT OUT  RESULTS
C
          WRITE(7,211)
211       FORMAT(///6X,1HX,12X,1HY,8X,8HX OR LNX,3X,11HPROBABILITY,
     1 4X,6HPROBIT,5X,5HDPROB/)
          DO 65 I=1,N
65        WRITE(7,212)X(I),Y(I),XZ(I),PR(I),PT(I),DPR(I)
212       FORMAT(F10.2,3X,F10.1,3X,F10.4,2XF10.5,2X,F10.4,2X,F10.5)
60        CALL CLOSE(5)
          GO TO 1
          END
```

APPENDIX V Input data file for Fortran program

```
PLATELET VOLUME DISTRIBUTION FOR RAT NO. 1 POST DIVE
11.
339075.
16.
8545.
21.
24432.
26.
49667.
31.
81915.
36.
116328.
41.
149879.
46.
180259.
51.
206747.
56.
229375.
61.
248784.
66.
264744.
71.
278218.
76.
289956.
81.
300032.
86.
309024.
91.
316741.
96.
323697.
101.
329312.
106.
335106.
```

APPENDIX VI Lognormal distribution analysis for platelet sample - teletype output.

```
MONITOR T4G

$A DK1 5/TT 4,6,7

$GLOAD


LOADS T9A
>←NORMD

*IF USING LINE PRINTER - TYPE 1 FOR YES, 0 FOR NO

0


*TYPE DATA FILE NAME

RAT1

*TYPE 1 FOR NORMAL, 0 FOR LOGNORMAL
0

PLATELET VOLUME DISTRIBUTION FOR RAT NO. 1 POST DIVE

LOGNORMAL DISTRIBUTION ANALYSIS

NO. OF OBSERVATIONS       19

TOTAL COUNT       339075.00

P = ALN(X) + B


NO. OF OBSERVATIONS USED IN ANALYSIS  19


A =     2.0878       B =      -2.8742

MEAN =      3.7714
STANDARD DEVIATION =      0.4790

MODE = EXP(M-S*S) =     34.5373
MEDIAN = EXP(M) =    43.4428
EXPECTATION VALUE = EXP(M+0.5S*S) =      48.7229

DIFFERENCE  (AZ + B) - PROBIT
   16.0 -0.128    21.0 -0.057    26.0 -0.020    31.0 -0.004    36.0  0.011
   41.0  0.025    46.0  0.040    51.0  0.057    56.0  0.073    61.0  0.085
   66.0  0.099    71.0  0.108    76.0  0.109    81.0  0.101    86.0  0.076
   91.0  0.036    96.0 -0.037   101.0 -0.161   106.0 -0.414
CORRELATION COEFFICIENT =   0.9941
```

```
*OMIT OBSERVATIONS - TYPE 1 FOR YES, 0 FOR NO
1
*HOW MANY AT START
0
*HOW MANY AT END
4


NO. OF OBSERVATIONS USED IN ANALYSIS  15


A =      1.9586       B =     -2.4187

MEAN =     3.7877
STANDARD DEVIATION =     0.5106

MODE = EXP(M-S*S) =    34.0211
MEDIAN = EXP(M) =    44.1525
EXPECTATION VALUE = EXP(M+0.5S*S) =      50.2990

DIFFERENCE  (AZ + B) - PROBIT
  16.0 -0.031    21.0  0.005    26.0  0.014    31.0  0.008    36.0  0.004
  41.0  0.000    46.0  0.001    51.0  0.004    56.0  0.008    61.0  0.009
  66.0  0.013    71.0  0.013    76.0  0.005    81.0 -0.011    86.0 -0.044
CORRELATION COEFFICIENT =    0.9999


*OMIT OBSERVATIONS - TYPE 1 FOR YES, 0 FOR NO
0
```

X	Y	X OR LNX	PROBABILITY	PROBIT	DPROB
16.00	8545.0	2.7726	0.02520	3.0430	-0.00179
21.00	24432.0	3.0445	0.07205	3.5391	0.00070
26.00	49667.0	3.2581	0.14648	3.9483	0.00333
31.00	81915.0	3.4340	0.24158	4.2991	0.00267
36.00	116328.0	3.5835	0.34307	4.5964	0.00158
41.00	149879.0	3.7136	0.44202	4.8545	0.00029
46.00	180259.0	3.8286	0.53162	5.0791	0.00039
51.00	206747.0	3.9318	0.60974	5.2782	0.00144
56.00	229375.0	4.0254	0.67647	5.4574	0.00276
61.00	248784.0	4.1109	0.73371	5.6237	0.00294
66.00	264744.0	4.1897	0.78078	5.7746	0.00369
71.00	278218.0	4.2627	0.82052	5.9173	0.00340
76.00	289956.0	4.3307	0.85514	6.0587	0.00114
81.00	300032.0	4.3944	0.88485	6.1997	-0.00216
86.00	309024.0	4.4543	0.91137	6.3495	-0.00717
91.00	316741.0	4.5109	0.93413	6.5076	-0.01244
96.00	323697.0	4.5643	0.95465	6.6921	-0.01874
101.00	329812.0	4.6151	0.97268	6.9222	-0.02522
106.00	335196.0	4.6634	0.98856	7.2759	-0.03170

```
*TYPE DATA FILE NAME
```

APPENDIX VII Normal distribution analysis for platelet sample - line printer version.

a) Teletype Printout

```
TS Z2B-A


MONITOR T4G

$A DKI 5/TT 4,6/LF 7

$GLOAD


LOADS T9A
>←NORMD

*IF USING LINE PRINTER - TYPE 1 FOR YES, 0 FOR NO

1



*TYPE DATA FILE NAME

RAT1

*TYPE 1 FOR NORMAL, 0 FOR LOGNORMAL
1

DIFFERENCE  (AZ + B) - PROBIT
   16.0  0.564    21.0  0.272    26.0  0.066    31.0 -0.080   36.0 -0.174
   41.0 -0.228    46.0 -0.248    51.0 -0.243    56.0 -0.218   61.0 -0.181
   66.0 -0.128    71.0 -0.066    76.0 -0.004    81.0  0.059   86.0  0.114
   91.0  0.160    96.0  0.179   101.0  0.153   106.0  0.003
CORRELATION COEFFICIENT =     0.9835


*OMIT OBSERVATIONS - TYPE 1 FOR YES, 0 FOR NO
0


*TYPE DATA FILE NAME

↑C
TS Z2B-A


MONITOR T4G

$
```

b) Line Printer Printout

PLATELET VOLUME DISTRIBUTION FOR RAT NO. 1 POST DIVE

NORMAL DISTRIBUTION ANALYSIS

NO. OF OBSERVATIONS 19

TOTAL COUNT 339075.00

P = AX + B

NO. OF OBSERVATIONS USED IN ANALYSIS 19

A = 0.0408 B = 2.9538

MEAN = 50.1447
STANDARD DEVIATION = 24.5059

DIFFERENCE (AZ + B) - PROBIT

16.0	0.564	21.0	0.272	26.0	0.066	31.0	-0.080	36.0	-0.174
41.0	-0.228	46.0	-0.248	51.0	-0.243	56.0	-0.218	61.0	-0.181
66.0	-0.128	71.0	-0.066	76.0	-0.004	81.0	0.059	86.0	0.114
91.0	0.160	96.0	0.179	101.0	0.153	106.0	0.003		

CORRELATION COEFFICIENT = 0.9835

X	Y	X OR LNX	PROBABILITY	PROBIT	DPROB
16.00	8545.0	16.0000	0.02520	3.0430	0.05655
21.00	24432.0	21.0000	0.07205	3.5391	0.04510
26.00	49667.0	26.0000	0.14648	3.9483	0.01576
31.00	81915.0	31.0000	0.24158	4.2991	-0.02425
36.00	116328.0	36.0000	0.34307	4.5964	-0.06116
41.00	149879.0	41.0000	0.44202	4.8545	-0.08752
46.00	180259.0	46.0000	0.53162	5.0791	-0.09878
51.00	206747.0	51.0000	0.60974	5.2782	-0.09581
56.00	229375.0	56.0000	0.67647	5.4574	-0.08205
61.00	248784.0	61.0000	0.73371	5.6237	-0.06262
66.00	264744.0	66.0000	0.78078	5.7746	-0.03961
71.00	278218.0	71.0000	0.82052	5.9173	-0.01790
76.00	289956.0	76.0000	0.85514	6.0587	-0.00083
81.00	300032.0	81.0000	0.88485	6.1997	0.01116
86.00	309024.0	86.0000	0.91137	6.3495	0.01692
91.00	316741.0	91.0000	0.93413	6.5076	0.01813
96.00	323697.0	96.0000	0.95465	6.6921	0.01469
101.00	329812.0	101.0000	0.97265	6.9222	0.00833
106.00	335196.0	106.0000	0.98856	7.2759	0.00010